dbSNP: Database for Short Genetic Variations

Catalog of nucleotide changes for human and other model organisms

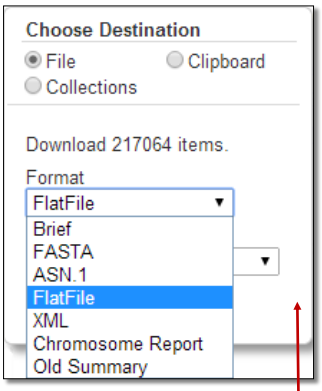
https://www.ncbi.nlm.nih.gov/snp/

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope and Access

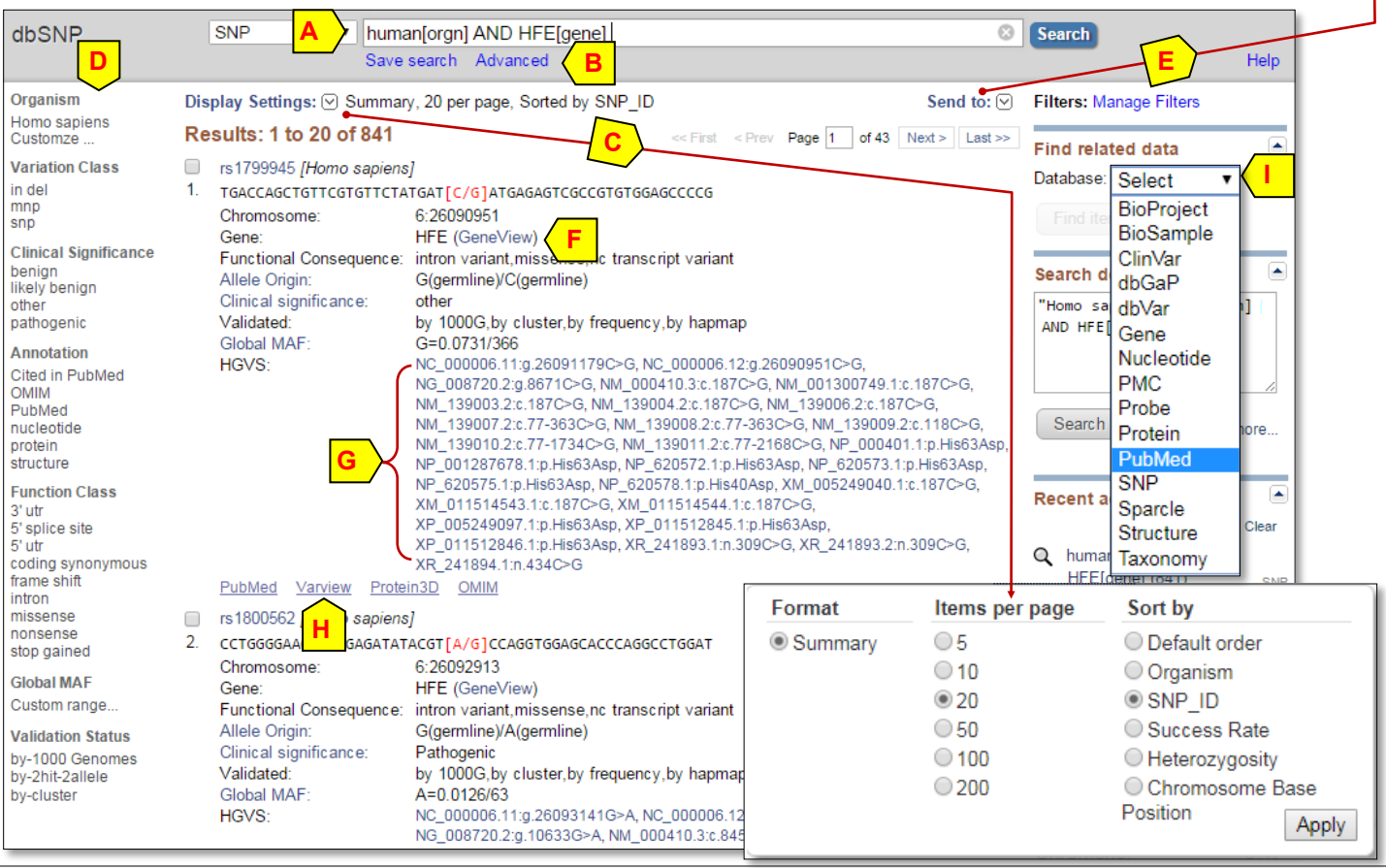
The NCBI Short Genetic Variations database (dbSNP [1]), commonly known as dbSNP, catalogs short variations in nucleotide sequences for human. These variations include single nucleotide variations, short nucleotide insertions and deletions, short tandem repeats. Short Genetic Variations may be common, thus representing true polymorphisms, or they may be rare. Some rare human entries have additional information associated with them, including disease associations (from ClinVar [2]), genotype information and allele origin, as some variations are somatic rather than germline events.

Short nucleotide variation data can be accessed via the SNP homepage and EUtils API:
www.ncbi.nlm.nih.gov/snp and www.ncbi.nlm.nih.gov/projects/SNP/SNPeutils.htm
VCF files and database bcp files are available for download through FTP and Aspera client at:
ftp.ncbi.nlm.nih.gov/snp/ and www.ncbi.nlm.nih.gov/public/?snp/organisms/
SNP data can also be accessed interactively through Variation Viewer:
www.ncbi.nlm.nih.gov/variation/view/
For information on how to find known SNPs in a sequence, refer to this handout:
ftp.ncbi.nlm.nih.gov/pub/factsheets/HowTo_Finding_SNP_by_BLAST.pdf



Searching for and Displaying SNP Records

You can search for variations on the dbSNP homepage by typing a query term in the search box and clicking the Search button (**A**). You can also use the Advanced (**B**) page to create complex queries to produce more precise results. The search below, "*hfe[gene] AND human[orgn]*", retrieves variations mapped to the human HFE gene. You can use options in the Display settings popup (**C**) to change the number of records displayed or sort retrieved variations in a different order. You can further narrow down retrieved variations by selecting filters present in the left column (**D**), or save them to a local file using the Send to (**E**) option. Use links to separate displays to see gene-centric listings (GeneView, **F**), graphical presentation under the context of genome or mRNA sequences (via HGVS names, **G**), or gene-centric display in a genomic context (Varview, **H**). Using the "Find related data" portet (**I**), you can retrieve related entries from other NCBI databases for the set of variations in the display.



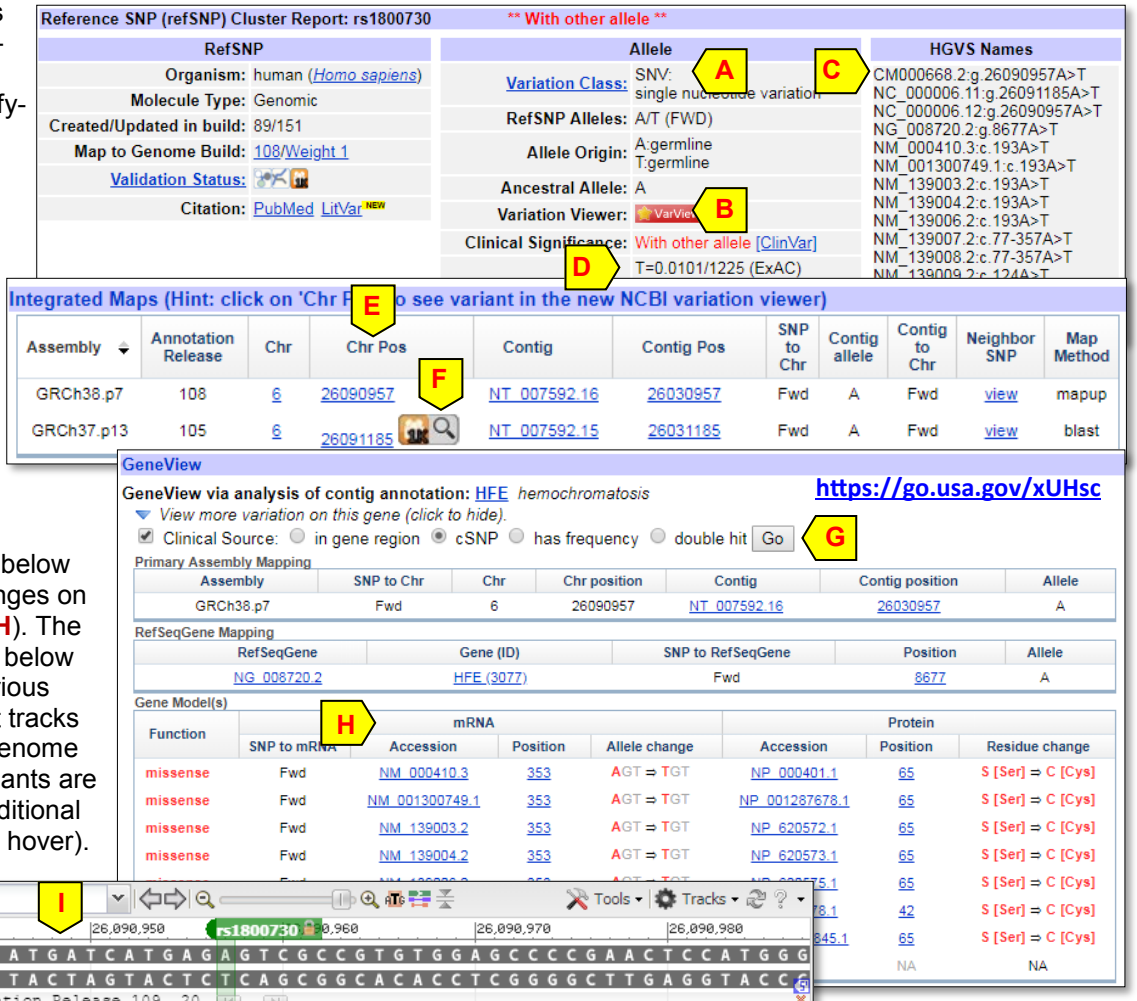
Contact: info@ncbi.nlm.nih.gov

The Reference SNP Cluster Report

The Reference SNP Cluster Report linked from rsID (rs1800730, shown in sections below and on p.3) provides details of a variation record. The report contains a summary of the allele (**A**), a link to the gene-centric display through the VarView icon (**B**, see p.4), mapping information in Human Genome Variation Society (HGVS) nomenclature (**C**), and minor allele frequencies (MAF, **D**) from various studies. The Integrated Maps table provides the genomic mapping details with the chromosomal coordinates (**E**) link to the same gene-centric display VarView icon provides. The magnifying glass (**F**) points to the 1000 Genomes Browser and provides genotyping details, if the variant this rsID represents is also called by that project.

For a summary of SNPs mapped to the gene, you can click the Go button (**G**) in the GeneView section to activate the SNP: GeneView display(p. 4).

The Gene Model(s) table below lists coordinates and changes on transcripts and proteins (**H**). The graphical panel (**I**) further below presents variants with various characteristics in different tracks (**J**) under the context of genome annotation. Individual variants are hyperlinked to provide additional details (**K**, activated upon hover).



The Submitter records table (**L**) lists alleles and flanking sequences from submitter SNPs (ssIDs) included in this reference SNP cluster. The ssIDs (**M**) link to submitter records with additional details.

Submitter records for this RefSNP Cluster (L)

The submission **ss244317424** has the longest flanking sequence of all cluster members and was used to instantiate sequence for **rs1800730** during BLAST analysis for the current build.

NCBI Assay ID	Handle|Submitter ID	Validation Status	ss to rs Orientation /Strand	Alleles	5' Near Seq 30 bp	3' Near Seq 30 bp	Entry Date	Update Date	Build Added	Molecule Type
ss2420855	HGBASE|SNP000007282		fwd/T	A/T	gctgttcgtgttctatgatcatgag	gtcgccgtgtggagccccgaactcc	11/07/00	10/10/03	89	Genomic
ss68969473	PERLEGEN|PGP04777602		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgccgtgtggagccccgaactccatggg	01/30/07	08/14/07	127	Genomic
ss160462894	ILLUMINA|HumanOmni1-Quad_v1-0_B_rs1800730...		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgccgtgtggagccccgaactccatggg	08/04/09	10/02/09	131	Genomic
ss233370756	1000GENOMES|pilot_1_CEU_2975385_chr6_261		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgccgtgtggagccccgaactccatggg	05/01/10	05/01/10	132	Genomic
ss244317424	|2010_April_001_075_HFE_235200_0003...		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgccgtgtggagccccgaactccatggg	06/16/10	06/16/10	132	Genomic
ss288288927	OMIM-CURATED-RECORDS|6335		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgccgtgtggagccccgaactccatggg	12/21/10	12/21/10	133	Genomic
ss342203118	NHLBI-ESP|ESP2500-chr6-26091185		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgccgtgtggagccccgaactccatggg	03/25/11	03/26/11	134	Genomic
ss410868034	ILLUMINA|Cardio-Metabo_Chip_A_chr6_26199164...		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgccgtgtggagccccgaactccatggg	06/07/11	06/07/11	135	Genomic
ss481067367	ILLUMINA|HumanOmni1-Quad_v1-0_rs1800730-128...		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgccgtgtggagccccgaactccatggg	01/30/12	01/31/12	137	Genomic
ss490921024	1000GENOMES|20110521_exome_428822_chr6_260...		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgccgtgtggagccccgaactccatggg	02/10/12	02/21/12	137	Genomic

M

The Reference SNP Cluster Report (cont.)

The FASTA Sequence section shows the sequences flanking the variation. It uses the exemplar submitter record to represent the variation by presenting the sequence as the 5'- (**A**) and 3'-(**B**) flanking sequences, with the allele (**C**) in the middle. The Population Diversity section (**D**) sums up available genotype and allele frequency information for various populations from different studies. More genotype details are available through the 1000 Genomes browser link (**E**).

Fasta sequence (Legend)

```
>gnl|dbSNP|rs1800730|allelePos=251|totalLen=501|taxid=9606|snpclass=1|alleles='A/T'|mol=Ge
 CAGGACTGCA ACTCACCCTT CACAAAATGA GGACCAGACA CAGCTGATGG TATGAGTTGA TGCAGGTGTG
 TGGAGCCTCA ACATCCTGCT CCCCTCCTAC TACACATGGT TAAGGCCTGT TGCTCTGTCT CCAGGTTCAC
 ACTCTCTGCA CTACCTCTTC ATGGGTGCCT CAGAGCAGGA CCTTGGTCTT TCCTTGTTTG AAGCTTTGGG
 CTACGT     GACCAGCTGT TCGTGTTCTA TGATCATGAG
 W
 GTCGCC     GGAGCCCCGA ACTCCATGGG TTTCCAGTAG AATTTCAAGC CAGATGTGGC TGCAGCTGAG
 TCAGAGTCTG AAAGGGTGGG ATCACATGTT CACTGTTGAC TTCTGGACTA TTATGGAAAA TCACAACCAC
 AGCAAGGGTA TGTGGAGAGG GGGCCTCACC TTCCTGAGGT TGTCAGAGCT TTTCATCTTT TCATGCATCT
 TGAAGGAAAC AGCTGGAAGT CTGAGGTCTT GTGGGAGCAG
```

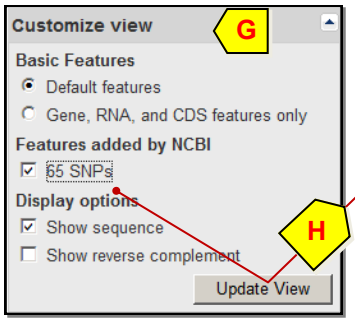
A B C D

Population Diversity (Alleles in RefSNP orientation) . See additional population frequency from 1000Genome [here] **E**

ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	Genotype Detail A/A	A/T	T/T	HWP	Alleles A	T
ss1319404492	EAS		1008	AF					1.00000000	
	EUR		1006	AF					0.98409998	0.01590000
	AFR		1322	AF					1.00000000	
	AMR		694	AF					0.99570000	0.00430000
	SAS		978	AF					0.99899995	0.00100000
ss1688187409	ExAc_Aggregated_Populations		121412	AF					0.98991036	0.01008961
ss233370756	pilot_1_CEU_low_coverage_panel		120	AF					0.96666664	0.03333334
ss342203118	ESP_Cohort_Populations		4552	GF	0.97803164	0.02196836		0.65472100	0.98901582	0.01098418
ss491881981	CSAgilent		1323	GF	0.98299998	0.01700000		1.00000000	0.99150002	0.00850000

Other Ways to Access Data from dbSNP

The SNP database is fully integrated with the Entrez system, enabling the access of variation data through links present in records from other NCBI databases. For example, you can use the SNP: GeneView or Variation Viewer link (**F**) found in the Related information section of a Gene database record to see a summary list of variations mapped to that gene. You can project variations mapped to a segment of the RefSeq genomic or a mRNA record (with NT_, NG_, NW_ or NM_ accessions) by using the Customize view (**G**) menu in the upper right hand corner of the sequence record, simply check the SNPs checkbox and click Update View (**H**) to activate the selection.




```
exon            1..236
                /gene="HFE"
                /gene_synonym="HFE1; HH; HLA-H; MVCD7; TFQTL2"
                /inference="alignment:Splign"
                /number=1
variation       34
                /gene="HFE"
                /gene_synonym="HFE1; HH; HLA-H; MVCD7; TFQTL2"
                /replace="c"
                /replace="g"
                /db_xref="dbSNP:62625316"
variation       49
                /gene="HFE"
                /gene_synonym="HFE1; HH; HLA-H; MVCD7; TFQTL2"
                /replace="c"
                /replace="t"
                /db_xref="dbSNP:62625317"
```

Related information
- Order cDNA clone
- 3D structures
- ■ ■ ■
- ClinVar
- Conserved Domains
- dbVar
- Full text in PMC
- Full text in PMC_nucleotide
- Functional Class
- GAP
- Gene neighbors
- Genome
- GEO Profiles
- GTR
- HomoloGene
- MedGen
- Nucleotide
- OMIM
- Probe
- Protein
- PubChem Compound
- PubChem Substance
- PubMed
- PubMed (GeneRIF)
- PubMed (OMIM)
- PubMed(nucleotide/PMC)
- RefSeq Proteins
- RefSeq RNAs
- RefSeqGene
- Related gene-specific medical variations
- SNP
- SNP: GeneView
- ■ ■ ■
- Variation Viewer

F

dbSNP also integrates disease-related nucleotide variations that were reported in literature and cited in rsID format, collected by OMIM, or submitted to ClinVar. The table below is the Allelic Variant display for OMIM record 613609, which cites the rsIDs in the dbSNP column (**I**).

I

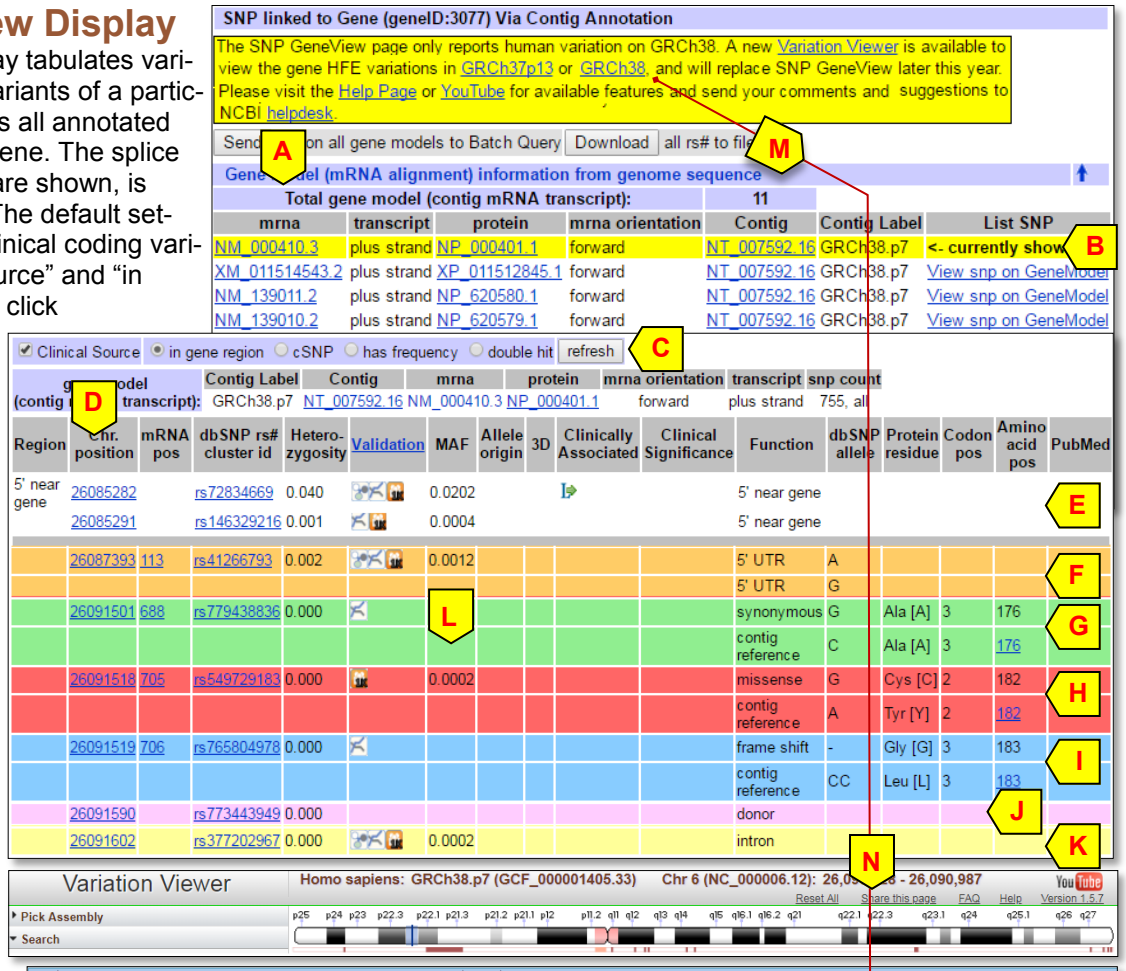
613609
HFE GENE; HFE
Allelic Variants (11 Selected Examples) :

All ClinVar Variants

Number	Phenotype	Mutation	dbSNP	ExAC	ClinVar
.0001	HEMOCHROMATOSIS, TYPE 1 PORPHYRIA CUTANEA TARDA, SUSCEPTIBILITY TO, INCLUDED PORPHYRIA VARIEGATA, SUSCEPTIBILITY TO, INCLUDED HEMOCHROMATOSIS, JUVENILE, DIGENIC, INCLUDED ALZHEIMER DISEASE, SUSCEPTIBILITY TO, INCLUDED TRANSFERRIN SERUM LEVEL QUANTITATIVE TRAIT LOCUS 2, INCLUDED MICROVASCULAR COMPLICATIONS OF DIABETES, SUSCEPTIBILITY TO, 7, INCLUDED	HFE, CYS282TYR	[rs1800562]	-	[RCV0002108
.0002	HEMOCHROMATOSIS, TYPE 1 MICROVASCULAR COMPLICATIONS OF DIABETES, SUSCEPTIBILITY TO, 7, INCLUDED	HFE, HIS63ASP	[rs1799945]	[rs1799945]	[RCV0000000
.0003	HEMOCHROMATOSIS, TYPE 1	HFE, SER65CYS	[rs1800730]	-	[RCV0002907
.0004	HFE INTRONIC POLYMORPHISM	HFE, 5569G-A	[rs1800758]	[rs1800758]	[RCV0000000
.0005	HFE POLYMORPHISM	HFE, VAL53MET	[rs28934889]	-	[RCV0000000
.0006	HFE POLYMORPHISM	HFE, VAL59MET	[rs111033557]	-	[RCV0000000
.0007	HEMOCHROMATOSIS, TYPE 1	HFE, GLN127HIS	[rs28934595]	-	[RCV0000000
.0008	HEMOCHROMATOSIS, TYPE 1	HFE, ARG330MET	[rs111033558]	-	[RCV0000000

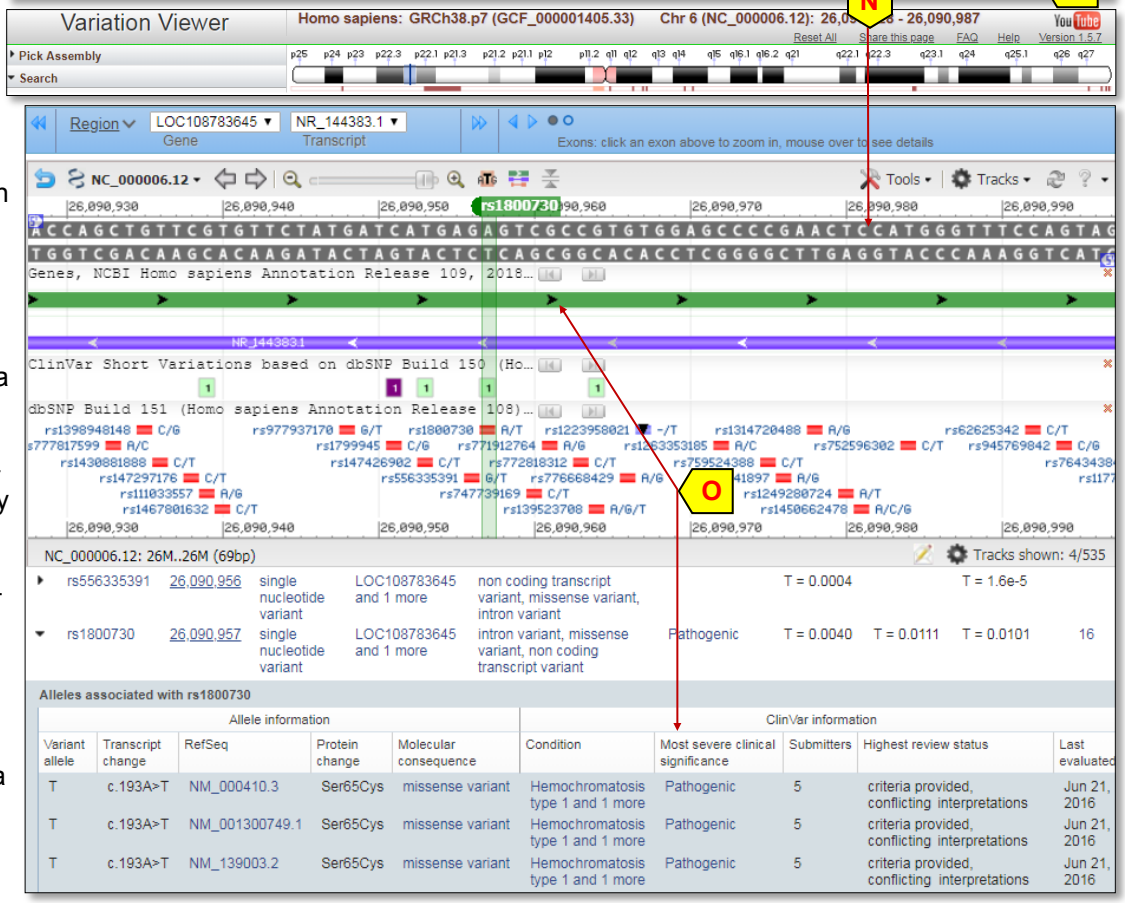
The SNP:GeneView Display

The SNP:GeneView display tabulates variations mapped to splice variants of a particular gene. At the top, it lists all annotated splice variants (**A**) of the gene. The splice variant, whose variations are shown, is highlighted in yellow (**B**). The default setting shows only the non-clinical coding variations. Check "Clinical Source" and "in gene region" options, then click "Reresh" (**C**) to see the complete list. The table arranges mapped variants by their chromosomal coordinates (**D**) and color-codes them by their function: **white** for "in gene region" (**E**), **orange** for UTR (**F**), **green** for synonymous (**G**), **red** for non-synonymous (**H**), **blue** for frame-shift (**I**), **purple** for splicing site (**J**), and **yellow** for intron (**K**). The MAF (**L**) column lists the global minor allele frequencies from the 1000 Genomes project.



Variation Viewer

The GeneView display contains a link (**M**) to an interactive display in the Variation Viewer (**N**), which can display variation mapping and molecular consequences within the GRCh37 or GRCh38 context for more focused examination - by correlating a variation and its molecular consequences in the data table with its genomic context in the graphical display (**O**). Filters in the left hand column (not shown) are available to selectively display variants of interest.



More information on this tool is available online as a video tutorial, in a factsheet, as well as in the more detailed online help:

Variation Viewer factsheet https://ftp.ncbi.nih.gov/pub/factsheets/Factsheet_Variation_Viewer.pdf
Online video tutorial https://www.youtube.com/watch?v=rnWZ9MFBwUM
Variation Viewer help https://www.ncbi.nlm.nih.gov/variation/view/help/